John von Neumann Institute for Computing

Simulation of Linker Histone-Chromatin Interactions

G. V. Pachov, R. R. Gabdoulline, R. C. Wade

published in

From Computational Biophysics to Systems Biology (CBSB07),
Proceedings of the NIC Workshop 2007,
Ulrich H. E. Hansmann, Jan Meinke, Sandipan Mohanty,
Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **36**, ISBN 978-3-9810843-2-0, pp. 69-74, 2007.

http://www.fz-juelich.de/nic-series/volume36

Simulation of Linker Histone-Chromatin Interactions

Georgi V. Pachov[1], **Razif R. Gabdoulline**[1,2] **and Rebecca C. Wade**[1]

[1] EML Research gGmbH, Molecular and Cellular Modeling Group,
Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany

[2] Center for Modelling and Simulation in the Biosciences (BIOMS),
Im Neuenheimer Feld 368, 69120 Heidelberg, Germany
E-mail: {*georgi.pachov, rebecca.wade*}@eml-r.villa-bosch.de

The dynamics and interactions between protein-protein and protein-DNA molecules appearing on different time and length scales in the cell are of fundamental interest. In chromatin, linker histone binds to the highly charged nucleosome facilitated by electrostatic attraction. However, its position and orientation with respect to the nucleosome core are unknown. Using an implicit representation of the solvent, rigid-body docking of a linker histone to the nucleosome has been performed by BD simulations. This reveals how the protein binds to DNA on the nucleosome. Two distinct binding sites on the linker histone have been identified by the docking simulation.

1 Introduction

In the cell nucleus, DNA wraps around histone proteins (forming nucleosome particles) and packs to a highly negatively charged structure, called chromatin fiber. The positioning of the nucleosomes within the fiber depends on the presence or lack of a linker histone protein[1]. However, the exact position and orientation of the linker histone with respect to the nucleosome particle is an unresolved problem, although some recent studies have proposed different binding modes[2–4].

Experimentally, it is known that the linker histone binds to the nucleosomal DNA[5] and to at least one of the linker DNAs. This, indicates the important role of the linker DNA's length in chromatin compaction[6]. Since the linker histone appears to be crucial for the opening and closing of the chromatin fiber during transcription and many other biological processes related to the linker histone and nucleosome in the higher order structure of the genome, we performed molecular docking simulations using a Brownian Dynamics (BD) algorithm[7]. This study aims to reveal the binding position and orientation of linker histone to the nucleosome.

2 Method

Using the crystal structure of the nucleosome core particle[8] (NCP, Protein Data Bank - PDB code 1kx5) and of the globular domain of the linker histone H5[9] (GH5, PDB code 1hst), we have performed molecular docking simulations with the Simulation of Diffusional Association (SDA) software package[10].

Chain B of the linker histone was used for the simulation. Since the histone tails and residues 119-128 in chain C are much floppier than the other parts of the nucleosome they were removed from it. In order to include linker DNAs, the tetranucleosome structure[12] was used. The DNAs from both nucleosome structures, the one without tails and one from

the tetranucleosome, were aligned with Pymol[13] and an additional 20 bp DNA was added to the nucleosome without tails. We will refer to this structure as tNCP. In addition, one of the linker DNAs in the tNCP was *shifted* away from the other linker DNA allowing more space for the linker histone to penetrate close to the nucleosomal DNA (see Fig. 1). This structure will be designated as sNCP. Partial charges and atomic radii were assigned with the PDB2PQR program[14] using the Amber force field. The electrostatic potential was computed for the NCPs and GH5 by solving the nonlinear Poisson-Boltzmann equation on grids with 257^3 and 200^3 points, respectively. The programs used were APBS[15] and UHBD[16]. The temperature was set to 300 K, the solvent dielectric constant to 78, the solute to 2 and the ionic strength to 100 mM.

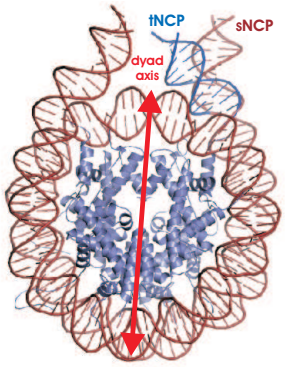

Figure 1. Difference between sNCP and tNCP structures. The linker DNA (ruby) in sNCP is shifted away in comparison with the linker DNA (blue) in tNCP. The red line shows the dyad axis of the nucleosome.

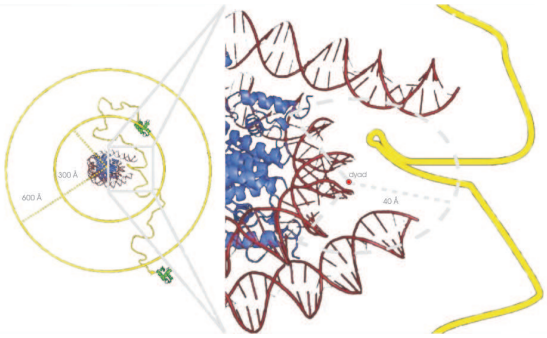

Figure 2. The picture shows the BD method, in which each trajectory of GH5 is started randomly on an inner sphere surface centered on the NCP and truncated on an outer sphere surface. The constraint region, where the docking solutions are recorded, is given in the zoomed picture. The red dot is the dyad point at the nucleosomal DNA.

To reduce the number of partial charge sites on the biomolecules, the ECM[17] program was applied. It derives effective charges fitted to reproduce the electrostatic potential in a uniform dielectric medium. On DNA, the charges were assigned to the *P* atoms only. The net formal charges are -237 e for the NCPs and 11 e for the GH5. An exclusion volume grid with 0.5 Å spacing was assigned to avoid van der Waals overlaps. The BD docking simulations were carried out with the SDA package modified in such a way that the docked complexes are recorded only if they satisfy predefined constraints (Fig. 2). For our case, we used two constraints: (i) a center-to-center between both particles ($\leq$ 73 Å) and (ii) dyad point-center distance between the dyad at the nucleosomal DNA and the center of the GH5 ($\leq$ 40 Å). In the simulations, the molecules are modeled as rigid bodies with the short-range attractive interactions neglected. The simulation method has been described in detail in the references[7,10,11]. The trajectories start at a center-to-center distance b=300 Å and finish at c= 600 Å (Fig. 2). The time step was set to 0.25 ps for center-to-center distances up to 130 Å and it increased linearly for larger distances. The interaction energies as well as the coordinates of the complexes satisfying the constraints were recorded.

For both sNCP and tNCP, five different runs with different random generators, i.e. different starting positions and orientations, were performed. For each nucleosome structure

25000 complexes were recorded and the 2500 lowest energy docked complexes were clustered with the program `clust`[18]. After clustering the representatives of 5 and 6 clusters were analyzed for the sNCP and tNCP, respectively.

3 Results and Discussion

The docking solutions for the linker histone GH5 were extensively investigated. Previous theoretical and experimental studies[2–4] identified the most important residues in binding, but the controversial results obtained left the problem open. Therefore, we aimed to find out which residues contribute to binding, how they influence the position and orientation of the linker histone with respect to the nucleosome and whether the distance between the linker DNAs, their mutual orientation and position affects the linker histone binding mode. The results are shown in Table 1 and Table 2 for sNCP and tNCP, respectively. The proximity of every residue considered as important to the nucleosome is given in two related ways as: (i) general - close to the nucleosomal DNA (nDNA), linker DNA (lDNA) or neither to the nDNA nor to the lDNA (Nn/lDNA) and (ii) average RMSD appearance $\bar{d}$ to the dyad point at the nucleosomal DNA (from all 5/6 representatives). The average interaction energy for each cluster varies between [-168;-104] kT and [-229;-128] kT for sNCP and tNCP, respectively.

Residue /	**nDNA** (N^{com})	**lDNA** (N^{com})	**Nn/lDNA** (N^{com})	$\bar{\mathbf{d}}$, Å
R42	3 (2357)	1 (72)	1 (68)	18.1
R94	3 (2357)	2 (140)	0 (0)	20.6
K97	3 (2357)	2 (140)	0 (0)	24.3
K85	2 (1981)	3 (516)	0 (0)	20.2
K40	1 (1676)	0 (0)	4 (821)	26.3
R47	1 (72)	3 (749)	1 (1676)	26.1
K69	1 (72)	3 (749)	1 (1676)	28.3
R73	1 (72)	3 (749)	1 (1676)	29.8
R74	1 (72)	2 (373)	2 (2052)	34.6
K52	2 (377)	2 (2052)	1 (68)	29.5
K55	1 (72)	2 (681)	2 (1744)	29.7
K59	0 (0)	1 (1676)	4 (821)	40.2

Table 1. Number of representatives of the clusters according to each residue located within 15 Å of the nucleosomal DNA (nDNA), linker DNAs (lDNA) and away (Nn/lDNA, $\leq$ 15 Å) from either nDNA or nDNA. All 2500 docked solutions (in brackets) are presented by 5 representatives. The distance $\bar{d}$ is the average RMSD from a specified residue (atom N_ζ for K and C_ζ for R) to the dyad point on the nucleosomal DNA for all representatives.

It is clearly seen for sNCP and tNCP that the site with residues Arg42, Arg94 and Lys97 appears to be closer to the nucleosomal DNA than to any of the linker DNAs (Fig. 3). Moreover, residue Arg42 takes the closest position to either nDNA or lDNA, i.e. less than 5 Å in comparison with the other residues[19]. On the other hand, the site with residues Arg47, Lys69, Arg73 and Arg74 shows a preference either for one of the linker DNA or for being away. This is in contrast to the theoretical model of Brown *et al.*[2]. The third distinct site in GH5 consisting of residues Lys52, Lys55 and Lys59 is mostly directed

Residue /	nDNA (N^{com})	lDNA (N^{com})	Nn/lDNA (N^{com})	$\bar{d}$, Å
R42	4 (1842)	2 (649)	0 (0)	24.6
R94	4 (1842)	2 (649)	0 (0)	24.5
K97	4 (1842)	1 (74)	1 (575)	24.3
K85	2 (480)	2 (1291)	2 (720)	27.5
K40	3 (554)	2 (1362)	1 (575)	22.7
R47	0 (0)	3 (1743)	3 (748)	31.46
K69	0 (0)	3 (1101)	3 (1390)	32.4
R73	0 (0)	3 (1101)	3 (1390)	33.7
R74	0 (0)	2 (526)	4 (1965)	33.7
K52	0 (0)	4 (1965)	2 (526)	27.3
K55	0 (0)	3 (1319)	3 (1172)	28.8
K59	1 (74)	4 (1965)	1 (452)	30.3

Table 2. The definition of the parameters is the same as in Table 1, but for the tNCP structure.

away from the NCP, although in the sNCP, Lys52 appears close to the nucleosome. Similar behaviour is observed for Lys40 as well. According to the docking data, it seems that the strongest contribution to binding close to the dyad point comes from residue Arg42 rather than residue Lys85 as stated by Fan and Roberts[3]. However, the general trend for the residues participating in the binding sites in both papers[2,3] is conserved in our simulations. It should be noted that there is a slight difference between the linker histone molecule used in the Brown *et al.*[2] model on one hand and in ours and in the Roberts model[3] on the other. In the former model, molecule A of the X-ray structure of GH5 resolved by Ramakrishnan *et al.*[9], has been used for homology modeling of H1[0] rather than molecule B used here and by Fan and Roberts[3]. The main difference is in the orientation of Lys85, which is directed to the Arg42 site in molecule B, while in molecule A it appears close to the helix 3 site. In addition, the relative orientation of Arg42 and Lys40 differs in both molecules by approximately 90°. Therefore, the binding sites identified in both studies do not differ

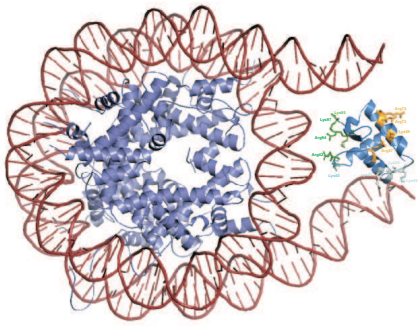

Figure 3. Representative of the largest (1676 solutions) and second lowest energy (-156 kT) cluster of the docked structures to sNCP.

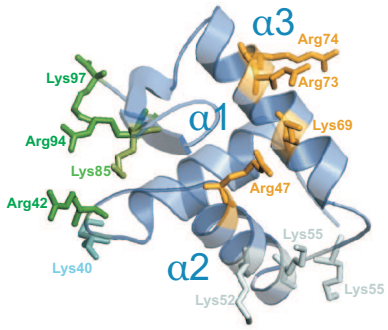

Figure 4. Linker histone structure with the analysed residues and binding sites according to Table 1 and Table 2. The orientation is the same as in Fig. 3

much from each other, when taking into account the above considerations. In fact, this is the reason why Fan and Roberts[3] consider Lys85 as a separate binding site while Brown *et al*.[2] assign Lys85 to the helix-3 binding site. Another difference in the final results comes from the assignment of Lys40 to a binding and nonbinding site in the Fan and Roberts[3] and the Brown *et al*.[2] models, respectively. The latter is deduced from a FRAP experiment performed with the $H1^0$ linker histone and is in agreement with our study only for the sNCP structure (see Table 1). However, the interpretation of the results for the tNCP has to be taken with care since this crystal structure does not include the linker histone. Apparently, due to the closeness of both linker DNAs in the tNCP structure, GH5 cannot penetrate to the nucleosomal DNA with the helix-3 site (see Table 2). On the other hand, both β sheets and the loop between helix-1 and helix-2, which include Arg42, Arg94, Lys97, Lys85 and Lys40 residues, need less space than helix-3 to fit in between the linker DNAs as indicated in Table 2. Therefore, from this point of view our results for sNCP seem more consistent with the experimental data[2].

4 Conclusion

The computational docking of a linker histone to the nucleosome has revealed two distinct binding sites on GH5 consisting of: (1) Arg42, Arg94, Lys97 with partial contribution of Lys85, which most probably bind to the nucleosomal DNA and (2) Arg47, Lys69, Arg73 and Arg74, which seem to contact one of the linker DNAs (see Fig. 4). The site including Lys52, Lys55 and Lys59 appears to be a nonbinding site with respect to the NCP, although some of the residues are located close to the sNCP in some of the solutions. In addition, the docking results clearly indicate the importance of the distance and orientation between both DNA arms, which might change the binding modes significantly. Therefore, including the flexibility in future investigations is necessary.

To obtain more insights into the exact contribution and position of each residue to binding, we aim to investigate the effects of single and double mutations in GH5 on binding. Moreover, simulation of diffusional motion of both molecules for computation of the association rates and encounter complex formation is under study. All this would be important for further development of models regarding the chromatin structure and compaction.

Acknowledgments

The work was supported by the German Research Foundation (DFG), the Center for Modelling and Simulation in the Biosciences (BIOMS) and the Klaus Tschira Foundation (KTF).

References

1. P. JJ. Robinson and D. Rhodes, *Structure of the '30 nm' chromatin fibre: A key role for the linker histone*, Curr. Opin. Str. Biol. **16**, 336–343, 2006.
2. D. T. Brown, T. Izard and T. Misteli, *Mapping the interaction surface of linker histone $H1^0$ with the nucleosome of native chromatin in vivo*, Nat. Str. Mol. Biol. **13**, 250–255, 2006.

3. L. Fan and V. A. Roberts, *Complex of linker histone H5 with the nucleosome and its implications for chromatin packing*, PNAS **103**, 8384–8389, 2006.
4. M. M. S. Bharath, N. R. Chandra and M. R. S. Rao, *Molecular modeling of the chromatosome particle*, Nucl. Acid Res. **31**, 4264–4274, 2003.
5. A. A. Travers, *The location of linker histone in the nucleosome*, Trends Biochem. Sci. **24**, 4–7, 1999.
6. P. JJ. Robinson and D. Rhodes, *EM measurements define the dimension of the "30-nm" chromatin fiber: Evidence for a compact, interdigitated structure*, PNAS **103**, 6506–6511, 2006.
7. D. L. Ermak and J. A. McCammon, *Brownian dynamics with hydrodynamic interactions*, J. Chem. Phys. **69**, 1352–1360, 1978.
8. C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder and T. J. Richmond, *Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å*, J. Mol. Biol. **319**, 1097–1113, 2002.
9. V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee and R. M. Sweet, *Crystal structure of globular domain of histone H5 and its implications for nucleosome binding*, Nature **362**, 219–223, 1993.
10. R. R. Gabdoulline, R. C. Wade, *Brownian dynamics simulation of protein-protein diffusional encounter*, Methods. **14**, 329–341, 1998.
`http://projects.villa-bosch.de/mcmsoft/sda/4.23/index.html`
11. R. R. Gabdoulline, R. C. Wade, *Protein-Protein association: Investigation of factors influencing association rates by brownian dynamics simulations*, J. Mol. Biol. **306**, 1139–1155, 2001.
12. T. Schalch, S. Duda, D. F. Sargent, T. J. Richmond, *X-ray structure of a tetranucleosome and its implications for the chromatin fibre*, Nature Lett. **436**, 138–141, 2005.
13. W. L. DeLano, *The Pymol Molecular Graphics System (2002)*.
`http://www.pymol.org`
14. T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, *PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations*, Nucl. Acids Res. **32**, W665–W667, 2004.
15. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon *Electrostatics of nanosystems: application to microtubules and the ribosome*, Proc. Natl. Acad. Sci. **98**, 10037–10041, 2001.
16. M. E. Davis, J. D. Madura, B. A. Luty, J. A. McCammon, *Electrostatic and diffusion of molecules in solution: simulations with the University-of-Houston- Brownian Dynamics program*, Comput. Phys. Commun. **62**, 187–197, 1991.
17. R. R. Gabdoulline, R. C. Wade, *Effective charges for macromolecules in solvent*, J. Phys. Chem. **100**, 3868–3878, 1996.
18. Unpublished software developed by Domantas Motiejunas (MCM Group, EML Research gGmbH).
19. Unpublished data.